STIC Search Report
Biotech-Chem Library

USPTO

TO: David Lamberston **Art Unit:** 1636 **Location:** rem/2B79/2C70 **Serial Number:** 10/042059 **Thursday, October 27, 2005**	**From:** Beverly Shears **Location:** Biotech-Chem Library REM 1A54 **Phone:** 571-272-2528 beverly.shears@uspto.gov

Search Notes

Protein Sequence Searches – February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.

- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (uniPARC) at:

  http://www.pir.uniprot.org/database/archive.shtml

If you have any questions regarding this information or your results, please contact any STIC searcher.

sira
Search and Information
Resources Administration

STIC-Biotech/ChemLib *CRFE* *168850*

From: Lambertson, David
Sent: Tuesday, October 18, 2005 11:06 AM
To: STIC-Biotech/ChemLib
Cc: Lambertson, David
Subject: Search Request

Search Request

Examiner's Name:	David Lambertson
Examiner #:	79514
Art Unit:	1636
Room #:	Remsen 02B79
Mailbox room#:	Remsen 02C70
Phone:	(571) 272-0771
Results Format:	paper

Serial # 10/042059

Please Search:

Nucleic Acid databases for:

SEQ ID NO: 1 *na 843*

Including:

1. Default Search. *me*

2. Interference Search.

Thanks,
Dave.

10/25/05
IH.

10/18/2005

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: October 20, 2005, 14:13:47 ; Search time 3558 Seconds
(without alignments)
9018.600 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 atggacttatgaagccaga.........tctccctggcttgtactag 843
Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	299	35.5	1185	9	CNS06T5P	AL414131 T7 end of
c	2	176.6	20.9	937	9	CNS06TV7	AL415049 T3 end of
c	3	176.6	20.9	1052	9	CNS06XEJ	AL419633 T3 end of
c	4	140	16.6	948	7	CO005472	CO005472
c	5	137	16.3	574	8	BZ303243	BZ303243 KD2284.p1
	6	132.8	15.8	738	7	CF877247	CF877247 tric075xb
	7	132.8	15.8	816	7	CB905545	CB905545 tric075xb
c	8	125.4	14.9	788	7	CO165801	CO165801 tric075xb
c	9	121.6	14.4	660	5	BX253798	BX253798
	10	116.8	13.9	798	5	CB619881	CB619881 OSIIEa04L
	11	116.8	13.8	652	6	CD423256	CD423256
c	12	113.2	13.5	747	9	CNS06TWI	CNS06TWI
c	13	113.2	13.4	710	7	CF475712	CF475712 RTWW2_11_
	14	111.2	13.2	1222	3	AY103806	AY103806 Zea mays
c	15	107.4	12.7	827	7	CO162312	CO162312 FLD1_34_B
	16	106	12.6	988	1	AJ833468	AJ833468
	17	105.8	12.6	602	6	CD982471	CD982471 QAXSc04.x
	18	105.8	12.5	903	7	CN149106	CN149106 WOUND1_60
	19	105.6	12.5	527	6	CO165812	CO165812 FLD1_57_F
c	20	104.6	12.4	606	4	BI995968	BI995968 103103zC0
c	21	103	12.2	737	7	CO005473	CO005473 EST793808
c	22	103	12.2	867	5	CB619882	CB619882 OSIIEa04L
c	23	102.8	12.2	727	6	CA121888	CA121888 SCJFLR105
	24	102.2	12.1	680	6	CA080410	CA080410 SCVPAM105
	25	100.8	12.0	551	5	BQ701847	BQ701847 NXSI_121
	26	99.4	11.8	605	7	CN913779	CN913779 030108ABM
	27	98.6	11.7	607	8	B2296833	B2296833 CG3070.r1
	28	98	11.6	495	1	AA786587	AA786587 mb08a1.f
c	29	97.8	11.6	821	7	CK196770	CK196770 FGAS00523
	30	97.4	11.5	837	7	CO366854	CO366854 RTK1_30_C
	31	97	11.5	596	8	B2298767	B2298767 CG4569.r1
	32	97	11.5	781	7	CF667838	CF667838 RTCNT1_32
c	33	96.6	11.5	853	7	CO366784	CO366784 RTK1_30_C
	34	96.6	11.4	657	9	AG273371	AG273371 Cyanidios
c	35	96	11.4	812	7	CN148636	CN148636 WOUND1_57
c	36	95	11.3	885	7	CV274873	CV274873 WS0174_B2
	37	94.8	11.2	857	7	CF635400	CF635400 zmrww00_c
	38	94.6	11.2	480	7	CV002646	CV002646 na003-13c
	39	94.4	11.2	707	6	CA239381	CA239381 SCRFPLS03
	40	94.2	11.2	785	7	CN523137	CN523137 GO015M01.
	41	93.6	11.1	552	4	BI718858	BI718858 1031032C0
	42	93.6	11.1	1265	3	CNS0A6WL	CNS0A6WL Arabidops
	43	93.6	11.1	1329	5	BX824460	BX824460 Arabidops
	44	93.2	11.1	698	2	BB577489	BB577489 L48-2335T
c	45	93.2	11.1	830	6	CB622596	CB622596 OSIIEa09I

ALIGNMENTS

RESULT 1
CNS06T5P

```
LOCUS       CNS06T5P        1185 bp    DNA    linear    GSS 05-JUL-2001
DEFINITION  T7 end of clone AW0AA029A11 of library AW0AA from strain CLIB 89 of
            Yarrowia lipolytica, genomic survey sequence.
ACCESSION   AL414131
VERSION     AL414131.1  GI:12186949
KEYWORDS    GSS.
SOURCE      Yarrowia lipolytica
ORGANISM    Yarrowia lipolytica
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE   1  (bases 1 to 1185)
AUTHORS     Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
TITLE       Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
JOURNAL     FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE     20584711
PUBMED      11152876
REFERENCE   2  (bases 1 to 1185)
AUTHORS     Casaregola,S., Neuveglise,C., Lepingle,A., Bon,E., Feynerol,C.,
            Artiguenave,F., Wincker,P. and Gaillardin,C.
TITLE       Genomic exploration of the hemiascomycetous yeasts: 17. Yarrowia
            lipolytica
JOURNAL     FEBS Lett. 487 (1), 95-100 (2000)
MEDLINE     20584727
PUBMED      11152892
REFERENCE   3  (bases 1 to 1185)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
```

```
OM nucleic - nucleic search, using sw model

                    Copyright (c) 1993 - 2005 Compugen Ltd.
                    GenCore version 5.1.6

Run on:             October 20, 2005, 15:44:38 ; Search time 900 Seconds
                    (without alignments)
                    7726.691 Million cell updates/sec

Title:              US-10-042-059B-1
Perfect score:      843
Sequence:           1 atggacttatgaagccaga..........tctcctcggcttgtactag 843

Scoring table:      IDENTITY NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           9772231 seqs, 4124568258 residues

Total number of hits satisfying chosen parameters:    19544462

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :          Published Applications NA:*
        1:  /cgn2_6/ptcdata/1/pubpna/US07_PUBCOMB.seq:*
        2:  /cgn2_6/ptcdata/1/pubpna/PCT_NEW_PUB.seq:*
        3:  /cgn2_6/ptcdata/1/pubpna/US06_NEW_PUB.seq:*
        4:  /cgn2_6/ptcdata/1/pubpna/US06_PUBCOMB.seq:*
        5:  /cgn2_6/ptcdata/1/pubpna/US07_NEW_PUB.seq:*
        6:  /cgn2_6/ptcdata/1/pubpna/PCTUS_PUBCOMB.seq:*
        7:  /cgn2_6/ptcdata/1/pubpna/US08_NEW_PUB.seq:*
        8:  /cgn2_6/ptcdata/1/pubpna/US08_PUBCOMB.seq:*
        9:  /cgn2_6/ptcdata/1/pubpna/US09A_PUBCOMB.seq:*
       10:  /cgn2_6/ptcdata/1/pubpna/US09B_PUBCOMB.seq:*
       11:  /cgn2_6/ptcdata/1/pubpna/US09C_PUBCOMB.seq:*
       12:  /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
       13:  /cgn2_6/ptcdata/1/pubpna/US0A_PUBCOMB.seq:*
       14:  /cgn2_6/ptcdata/1/pubpna/US0B_PUBCOMB.seq:*
       15:  /cgn2_6/ptcdata/1/pubpna/US0C_PUBCOMB.seq:*
       16:  /cgn2_6/ptcdata/1/pubpna/US0D_PUBCOMB.seq:*
       17:  /cgn2_6/ptcdata/1/pubpna/US0E_PUBCOMB.seq:*
       18:  /cgn2_6/ptcdata/1/pubpna/US0F_PUBCOMB.seq:*
       19:  /cgn2_6/ptcdata/1/pubpna/US0G_PUBCOMB.seq:*
       20:  /cgn2_6/ptcdata/1/pubpna/US0H_PUBCOMB.seq:*
       21:  /cgn2_6/ptcdata/1/pubpna/US0I_PUBCOMB.seq:*
       22:  /cgn2_6/ptcdata/1/pubpna/US10_NEW_PUB.seq:*
       23:  /cgn2_6/ptcdata/1/pubpna/US11A_PUBCOMB.seq:*
       24:  /cgn2_6/ptcdata/1/pubpna/US11_NEW_PUB.seq:*
       25:  /cgn2_6/ptcdata/1/pubpna/US60_NEW_PUB.seq:*
       26:  /cgn2_6/ptcdata/1/pubpna/US60_PUBCOMB.seq:*
```

SUMMARIES

```
                                                  %
Result                     Query
No.     Score    Match    Length   DB   ID              Description
------  -------  -------  -------  ---  ------------------------------
  1      843     100.0     843     13   US-10-042-059A-1   Sequence 1, Appli
  2      843     100.0    1655     13   US-10-042-059A-3   Sequence 3, Appli
  3      270      32.0     807     21   US-10-741-849-6032 Sequence 6032, Ap
c 4      156.4    18.6     869     20   US-10-472-587-26   Sequence 26, Appl
c 5      141.2    16.7    2920     15   US-10-128-714-2249 Sequence 249, App
c 6      141.2    16.7    2987     15   US-10-128-714-5249 Sequence 5249, Ap
c 7      131.4    15.6    1016     19   US-10-437-963-93956 Sequence 93956, A
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
        8:  131.4   15.6  1231   9   US-09-454-279-13       Sequence 13, Appl
        9:  131.4   15.6  1231  17   US-10-624-061-13       Sequence 13, Appl
       10:  118.2   14.0  1423  18   US-10-224-599-123820   Sequence 123820,
       11:  116.8   13.9   658  20   US-10-653-047-6992     Sequence 6992, Ap
       12:  116.6   13.8  1146  18   US-10-425-114-7285     Sequence 7285, Ap
       13:  116     13.8  1376  20   US-10-425-115-102347   Sequence 102347,
       14:  112.2   13.3  1548  20   US-10-739-930-4455     Sequence 4455, Ap
       15:  112.2   13.3  1223  20   US-09-454-279-11       Sequence 11, Appl
       16:  111.2   13.2  1223  17   US-10-624-061-11       Sequence 11, Appl
       17:  111.2   13.2  1500  19   US-10-767-701-14353    Sequence 14353, A
       18:  110.6   13.1  1010  20   US-10-425-115-102351   Sequence 102351,
c      19:  110.6   13.1  1397  20   US-10-425-115-102346   Sequence 102346,
       20:  104.4   12.4  1217  14   US-10-425-114-7639     Sequence 7639, Ap
       21:  104.4   12.4  1217  18   US-10-425-114-7639     Sequence 7639, Ap
       22:  103.2   12.2  1368  20   US-09-454-279-15       Sequence 15, Appl
       23:  99.8    11.8  1020  17   US-10-624-061-15       Sequence 15, Appl
       24:  99.8    11.8  1020  18   US-10-425-115-102349   Sequence 102349,
       25:  99.4    11.8  1079   9   US-09-938-842A-972     Sequence 972, App
       26:  97.4    11.6   798  18   US-09-938-842A-972     Sequence 972, App
       27:  97.4    11.6   798  11   US-10-267-763-3        Sequence 3, Appli
c      28:  97.4    11.6  1006  14   US-10-267-763-3        Sequence 3, Appli
       29:  97.4    11.6  1006  17   US-10-267-763-9        Sequence 9, Appli
       30:  97      11.6  1430  20   US-10-739-930-4454     Sequence 4454, Ap
       31:  93.6    11.5  1005   9   US-09-938-842A-737     Sequence 737, App
       32:  93.6    11.1  1005  18   US-09-938-842A-737     Sequence 737, Appli
c      33:  93.6    11.1  1207  14   US-10-267-763-2        Sequence 2, Appli
       34:  93.6    11.1  1207  11   US-10-267-763-8        Sequence 8, Appli
       35:  93.6    11.1  1299   9   US-09-938-842A-972     Sequence 972, App
       36:  87.6    10.4  1653  18   US-10-424-599-93036    Sequence 93036, A
c      37:  81       9.6   880  18   US-10-425-114-3898     Sequence 3898, Ap
       38:  80       9.5  1314  20   US-10-425-115-151470   Sequence 151470,
       39:  78.4     9.3  1197  18   US-10-425-114-27576    Sequence 27576, A
       40:  77.2     9.2   524  20   US-10-425-115-102342   Sequence 102342,
       41:  76.2     9.0   768  19   US-10-437-963-49372    Sequence 49372, A
       42:  74.8     8.9  2154  19   US-10-437-963-83557    Sequence 83557, A
c      43:  74.2     8.8  1066  18   US-10-425-114-1332     Sequence 1332, Ap
       44:  74.2     8.8  1083  20   US-10-425-115-115559   Sequence 115559,
       45:  73       8.7   580  19   US-10-021-323-12291    Sequence 12291, A
```

ALIGNMENTS

```
RESULT 1
US-10-042-059A-1
; Sequence 1, Application US/10042059A
; Publication No. US2002019770A1
; GENERAL INFORMATION:
; APPLICANT: Rhein Biotech Gesellschaft fur neue biotechnologische Prozesse und
; APPLICANT: Produkte mbH
; TITLE OF INVENTION: Nucleic Acid Molecule Containing a Nucleic Acid Coding for a Poly
; TITLE OF INVENTION: with Chorismate Mutase Activity
; FILE REFERENCE: PA305580US-019
; CURRENT APPLICATION NUMBER: US/10/042,059A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: DE 199 19 124.7
; PRIOR FILING DATE: 1999-04-27
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 843
;   TYPE: DNA
;   ORGANISM: Hansenula polymorpha
US-10-042-059A-1

  Query Match     100.0%;   Score 843;  DB 13;   Length 843;
  Best Local Similarity   100.0%;
  Matches  843;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGGACTTTATGAAGCCAGAAAACCAGTGCTCGGACCTTGGCAACATTAGAGATGCCTTGGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGACTTTATGAAGCCAGAAAACCAGTGCTCGGACCTTGGCAACATTAGAGATGCCTTGGTC 60
```

OM nucleic - nucleic search, using sw model

Run on: October 20, 2005, 14:14:27 ; Search time 198 Seconds
                                             (without alignments)
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2005 Compugen Ltd.
                                             6966.572 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 atggacttatgaagccaga........tcccctcctggcttgtactag 843

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
                1: /cgn2_6/ptcdata/1/ina/5A_COMB.seq:*
                2: /cgn2_6/ptcdata/1/ina/5B_COMB.seq:*
                3: /cgn2_6/ptcdata/1/ina/6A_COMB.seq:*
                4: /cgn2_6/ptcdata/1/ina/6B_COMB.seq:*
                5: /cgn2_6/ptcdata/1/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/ptcdata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	270	32.0	948	4	US-09-248-796A-4110	Sequence 4110, Ap
2	131.4	15.6	1231	4	US-09-454-279-11	Sequence 11, Appli
3	111.2	13.2	1223	4	US-09-454-279-13	Sequence 13, Appli
4	104.4	12.4	1217	3	US-09-610-040-4	Sequence 4, Appli
5	104.4	12.4	1217	3	US-09-610-040-10	Sequence 10, Appli
6	104.4	12.4	1217	4	US-10-267-763-4	Sequence 4, Appli
7	104.4	12.4	1217	3	US-10-267-763-10	Sequence 10, Appl
8	99.8	11.8	1020	4	US-09-454-279-15	Sequence 15, Appl
9	97.4	11.6	1006	3	US-09-610-040-9	Sequence 9, Appl
10	97.4	11.6	1006	4	US-10-267-763-9	Sequence 9, Appli
11	97.4	11.6	1006	3	US-10-267-763-3	Sequence 3, Appli
12	97.4	11.6	1006	4	US-10-267-763-9	Sequence 9, Appli
13	93.6	11.1	1207	3	US-09-610-040-2	Sequence 2, Appli
14	93.6	11.1	1207	3	US-09-610-040-8	Sequence 8, Appli
15	93.6	11.1	1207	4	US-10-267-763-2	Sequence 2, Appli
16	93.6	11.1	1207	4	US-10-267-763-8	Sequence 8, Appli
17	61	7.2	780	4	US-09-454-279-17	Sequence 17, Appl
18	60.8	7.2	7218	1	US-08-233-463-14	Sequence 14, Appl
19	59.8	7.1	579	4	US-09-454-279-7	Sequence 7, Appl
20	53.4	6.3	525	4	US-09-454-279-5	Sequence 5, Appli
21	37.6	4.5	4079	4	US-09-016-434-1248	Sequence 1248, Ap
22	37.6	4.5	4517	4	US-09-949-016-4573	Sequence 4573, Ap
23	37.6	4.5	4519	4	US-09-023-655-1202	Sequence 1202, Ap
24	37.6	4.5	4519	4	US-09-949-016-650	Sequence 650, App
25	37.6	4.5	4519	3	US-09-492-027-3	Sequence 3, Appli
26	37.2	4.4	1590	4	US-09-252-991A-15307	Sequence 15307, A
27	37.2	4.4	2352	4	US-09-252-991A-15532	Sequence 15532, A
28	37.2	4.4	2916	4	US-09-252-991A-15259	Sequence 15259, A
29	36.8	4.4	2169	3	US-09-103-840A-2	Sequence 2, Appli
30	36.6	4.3	5176	3	US-09-434-408-3	Sequence 3, Appli
31	36.6	4.3	5176	4	US-09-610-040-6	Sequence 6, Appl
32	36.4	4.3	601	4	US-10-267-763-6	Sequence 6, Appl
33	35.4	4.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
34	35.4	4.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
35	35	4.2	2097	4	US-09-252-991A-13635	Sequence 13635, A
36	35	4.2	2331	4	US-09-252-991A-13809	Sequence 13809, A
37	34.6	4.1	789	4	US-09-252-991A-12701	Sequence 12701, A
38	34.6	4.1	1632	4	US-09-252-991A-13259	Sequence 13259, A
39	34.4	4.1	3008	4	US-09-435-376-6	Sequence 6, Appli
40	34.2	4.1	1509	3	US-09-724-797-89	Sequence 89, Appl
41	34.2	4.1	1509	4	US-09-252-991A-13259	Sequence 13259, A
42	34.2	4.1	11958	3	US-09-134-246-8	Sequence 8, Appli
43	34	4.0	11958	4	US-09-664-186-8	Sequence 8, Appli
44	34	4.0	705	4	US-09-270-767-5061	Sequence 5061, Ap
45	34	4.0	978	4	US-09-270-767-20343	Sequence 20343, A
					US-09-902-540-6442	Sequence 6442, Ap

ALIGNMENTS

RESULT 1
US-09-248-796A-4110
; Sequence 4110, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 4110
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-4110

  Query Match   32.0%; Score 270; DB 4; Length 948;
  Best Local Similarity   59.1%; Pred. No. 1.4e-72;
  Matches   462; Conservative   0; Mismatches 320; Indels   0; Gaps   0;

Qy   1 ATGGACTTTATGAAGCCAGAAAACAGTGCTGGACCTTGGCAACATTAGAGATGCCTTGGTC 60
          ||| || ||||| || || || ||| ||| || ||| |||| || | ||| |||||
Db   142 ATGGATTTTATGAAACCAGAAACCTGTGCCTTGATCTTGCCAACATCCGTCAAGCANTTGGTTA 201

Qy   61 CGGATGGGAGGATACGATCATCTTCAACTTTATCGAGCGGTCGCAGTTCTATGCGTCGCGCC 120
          ||| ||||| ||||| ||| ||| ||||| | || | | |||| |||| | ||
Db   202 AGGATGGGAAGATACTATTGTGTTTGATTTTAAATGGAAAGATCTCAAATTGTTTTCAGTTCCCCA 261

Qy   121 TCGGTATACAAAGTCAACCAGTTCCCTATTCCCAAACTTCGGACGGCTCGTTCTTCGGACTCG 180
          || ||| ||||| ||| ||| |||||| || ||||| |||| | | | |||||| | || |
Db   262 TCAGTTTATGAAAGAAATAAAATATATATATATCCCAATTTTGATGGACTTTTTTGGAATGG 321

Qy   181 CTGTTGTTGCACGACGACGAATCCATTCGCAGGTGAGGAGTACGACGCGCCAGACGAG 240
          || ||| | || |||||| ||| | ||| || ||| || ||| | | |||||||||
Db   322 GCTTTGTTACAATGGAAGTTGCTCATTCTCAAATCAGACGTTATGAAGGCACACCAGACGAA 381

Qy   241 GTGGCTTTTTTCCCCAACGTGCTGGAAAAAAACGTTTCTGCCCAAGATCAACTACCCATCG 300
          || ||||||| || || ||| | ||||| ||| | | || | |||| |||| | | |
Db   382 ACTCCATTTTTTCCAGATCAATTGAAAAACTCCAATTTTTACCGGCCAAATCAATTATCCTAAA 441

Qy   301 GTGCTAGCCTCCTTACGCGCGGATGAAAGAAATCAACGTCAACAAAAGAGAGATTACTACACG 360
          || | | | || | | | | ||||| |||| | || ||| || | |||| | | | |
Db   442 ATATTGGGCCAAAATATTCTGATGAAAATTAATGTTAATTCTGAAAATAATGAAGTTTTATGTT 501

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2005 Compugen Ltd.

GenCore version 5.1.6

Run on: October 20, 2005, 14:09:57 ; Search time 591 Seconds (without alignments)
8443.895 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 843
Sequence: 1 atggacttatgaagccaga..........tctcctctggcttgtactag 843

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1:0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :
1: N_Geneseq_16Dec04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2002bs:*
9: geneseqn2003as:*
10: geneseqn2003bs:*
11: geneseqn2003cs:*
12: geneseqn2003ds:*
13: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	843	100.0	843	3	AAC81949	Aac81949 H. polymo
2	843	100.0	1655	3	AAC81950	Aac81950 H. polymo
3	270	32.0	807	12	ADP98547	Adp98547 C. albica
4	182.8	21.7	869	13	ADR85510	Adr85510 Aspergill
5	156.4	18.6	869	8	AAL50202	Aal50202 M sterili
c 6	141.2	16.7	989	13	ADR84923	Adr84923 Aspergill
c 7	141.2	16.7	2920	8	ABT17891	Abt17891 Aspergill
c 8	141.2	16.7	2987	8	ABT19705	Abt19705 Aspergill
c 9	141.2	16.7	6989	13	ADR84336	Adr84336 Aspergill
10	131.4	15.6	1231	10	ABX93057	Abx93057 Rice chor
11	131.4	15.6	1231	12	ADJ55263	Adj55263 Rice chor
12	116.8	13.9	658	3	AAF14469	Aaf14469 Aspergill
13	111.2	13.2	1223	10	ABX93056	Abx93056 Corn chor
14	111.2	13.2	1223	12	ADJ55261	Adj55261 Corn chor
c 15	104.4	12.4	1217	6	ABA91395	Aba91395 Arabidops
c 16	104.4	12.4	1217	6	ABA91401	Aba91401 Arabidops
17	99.8	11.8	1020	10	ABX93058	Abx93058 Soybean c
18	99.8	11.8	1020	12	ADJ55265	Adj55265 Soybean c
19	97.4	11.6	798	6	ABZ13167	Abz13167 Arabidops
20	97.4	11.6	993	6	ABA99614	Aba99614 Construct
21	97.4	11.6	1006	6	ABA99613	Aba99613 A. thalia
22	97.4	11.6	1006	6	ABA91400	Aba91400 Arabidops
23	97.4	11.6	1006	6	ABA91394	Aba91394 Arabidops
24	93.6	11.1	1005	6	ABZ12932	Abz12932 Arabidops
25	93.6	11.1	1207	6	ABA91393	Aba91393 Arabidops
c 26	93.6	11.1	1207	6	ABA91399	Aba91399 Arabidops
27	93.6	11.1	1274	3	AAC33324	Aac33324 Arabidops
28	73	8.7	580	13	ACN57510	Acn57510 Cotton gy
29	69.8	8.3	622	13	ACN50132	Acn50132 Cotton no
30	64.2	7.6	600	13	ADR64815	Adr64815 Cotton cD
31	61.8	7.3	258	6	ABZ71526	Abz71526 Corn tass
32	61	7.2	579	12	ADJ55257	Adj55257 Wheat cho
33	61	7.2	780	10	ABX93059	Abx93059 Wheat cho
34	59.8	7.1	780	12	ADJ55267	Adj55267 Wheat cho
35	59.8	7.1	579	10	ABX93054	Abx93054 Wheat cho
36	56.2	6.7	584	13	ACN50205	Acn50205 Cotton no
37	54.8	6.5	542	13	ACN57601	Acn57601 Cotton gy
38	53.4	6.3	525	12	ADJ55255	Adj55255 Soybean c
39	53.1	6.3	525	10	ABX93053	Abx93053 Soybean c
40	52.6	6.3	524	13	ADR64235	Adr64235 Cotton CD
41	49.4	5.9	2000	8	ADA71938	Ada71938 Rice gene
42	48.8	5.8	864	12	ADJ44958	Adj44958 Plant cDN
43	47.6	5.8	1379	6	ABA91794	Aba91794 Yeast mit
c 44	47	5.6	5858	8	AAD55726	Aad55726 Nephila m
c 45	46.8	5.6	180385	10	ADI13931	Adi13931 Osteoearth

ALIGNMENTS

RESULT 1
AAC81949
AAC81949 standard; DNA; 843 BP.

ID   AAC81949 standard; DNA; 843 BP.
XX
AC   AAC81949;
XX
DT   15-SEP-2003 (revised)
DT   01-MAR-2001 (first entry)
XX
DE   H. polymorpha chorismate mutase DNA.
XX
KW   Chorismate mutase; prephenate; selection marker; auxotrophic yeast; ds.
XX
OS   Pichia angusta.
XX
XX
XX
FH
PN   WO200065071-A1.
XX
PD   02-NOV-2000.
XX
XX
PF   27-APR-2000; 2000WO-EP003844.
XX
PR   27-APR-1999; 99DE-0119124.
XX
PA   (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
XX
PI   Gellissen G, Braus G, Pries R, Krappmann S, Strasser AW;
XX
DR   WPI; 2000-687355/67.
DR   P-PSDB; AAB11451.
XX
PT   Nucleic acids encoding chorismate mutase, useful for preparing an
PT   auxotrophic selection system for recombinant yeast and for recombinant
PT   protein expression.
XX
PS   Claim 1a; Page 57; 63pp; German.
XX
CC   This invention describes novel nucleic acids (I) that encode a
CC   polypeptide (II) with chorismate mutase (CM) activity (or its
CC   complementary strand). CM catalyzes conversion of chorismate to
CC   prephenate, an essential precursor for Phe and Tyr. (I) is a selection
CC   marker for construction of corresponding auxotrophic yeast (requiring Phe
CC   and Tyr) that are used for recombinant production of proteins. (I) allows
CC   selection of transformed yeast on simple media. (Updated on 15-SEP-2003

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

Run on: October 20, 2005, 14:12:47 ; Search time 4117 Seconds
(without alignments)
9921.726 Million cell updates/sec

Title: US-10-042-059B-1
Perfect score: 843
Sequence: 1 acggacttatgaagccaga.........tctcctctggctctgtactag 843

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2427607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	DB Length	DB ID	Description
	1	843	100.0	843	6	BD266599	BD266599 Nucleic a
	2	843	100.0	843	6	AX043758	AX043758 Sequence
	3	843	100.0	1652	6	AF204738	AF204738 Pichia an
	4	843	100.0	1655	6	BD266600	BD266600 Nucleic a
	5	843	100.0	1655	6	AX043760	AX043760 Sequence
c	6	325.2	38.6	110000	6	CR382131_20	CR382131_20 Continuation (21 o
c	7	281.8	33.4	110000	6	CR382138_20	CR382138_20 Continuation (21 o
c	8	270	32.0	948	6	AR548979	AR548979 Sequence
c	9	256.4	30.4	110000	8	AE016817_01	AE016817_01 Continuation (2 of
	10	217.4	25.8	2059	8	YSCARO7A	YSCARO7A Continuation
c	11	217.4	25.8	43376	8	Z49219	Z49219 S.cerevisia
c	12	217.4	25.8	165536	8	Z71255	Z71255 S.cerevisia
c	13	217.2	25.8	771	8	SCHRXVI	SCHRXVI Saccharom
c	14	210.4	25.0	110000	8	AY693179	AY693179 Saccharon
	15	189	22.4	110000	8	CR382122_07	CR382122_07 Continuation (12 o
	16	156.4	18.6	869	6	BD178335	BD178335 Transform
	17	156.4	18.6	869	8	AB116236	AB116236 Rosellini
	18	141.8	16.8	1348	8	AK105512	AK105512 Oryza sat
	19	131.4	15.6	1231	6	AR404687	AR404687 Sequence
	20	131.4	15.6	1250	8	AK068983	AK068983 Oryza sat
	21	113.6	13.5	3167	8	AF133241	AF133241 Emericell
	22	111.2	13.2	1223	6	AR404686	AR404686 Sequence
	23	110	13.0	1000	8	L47356	L47356 Lycopersico
	24	107.6	12.8	951	8	BT005306	BT005306 Arabidops
	25	107.6	12.8	1143	8	AK117860	AK117860 Arabidops
	26	104.4	12.4	1217	6	AR236635	AR236635 Sequence
c	27	104.4	12.4	1217	6	AR236641	AR236641 Sequence
c	28	104.4	12.4	1217	6	AX464576	AX464576 Sequence
c	29	104.4	12.4	1217	6	AX464582	AX464582 Sequence
	30	104.4	12.4	1217	8	AF131219	AF131219 Arabidops
	31	100.6	11.9	1100000	2	AP006499_05	AP006499_05 Continuation (6 of
	32	99.8	11.8	1020	8	AR404688	AR404688 Sequence
	33	97.4	11.6	798	6	AX506277	AX506277 Sequence
	34	97.4	11.6	829	8	AY133840	AY133840 Arabidops
	35	97.4	11.6	993	6	AX343935	AX343935 Sequence
	36	97.4	11.6	1006	6	AR236634	AR236634 Sequence
c	37	97.4	11.6	1006	6	AR236640	AR236640 Sequence
c	38	97.4	11.6	1006	6	AX343933	AX343933 Sequence
c	39	97.4	11.6	1006	6	AX464575	AX464575 Sequence
c	40	97.4	11.6	1006	8	AX464581	AX464581 Sequence
c	41	97.4	11.6	1006	8	L47355	L47355 Arabidopsis
c	42	97.4	11.6	1055	8	AY065238	AY065238 Arabidops
	43	93.6	11.1	1055	8	AX506042	AX506042 Sequence
c	44	93.6	11.1	1207	6	AR236633	AR236633 Sequence
c	45	93.6	11.1	1207	6	AR236639	AR236639 Sequence

ALIGNMENTS

RESULT 1
BD266599
LOCUS   BD266599   843 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION   Nucleic acid molecule, containing a nucleic acid which codes for a
             polypeptide with chorismate mutase activity.
ACCESSION   BD266599
VERSION   BD266599.1 GI:33076367
KEYWORDS   JP 2002542788-A/1.
SOURCE   Pichia angusta
ORGANISM   Pichia angusta
             Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
             Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE   1 (bases 1 to 843)
AUTHORS   Gellissen,G., Braus,G., Pries,R., Krappmann,S. and Strasser,A.W.
TITLE   Nucleic acid molecule, containing a nucleic acid which codes for a
             polypeptide with chorismate mutase activity
JOURNAL   Patent: JP 2002542788-A 1 17-DEC-2002;
             RHEIN BIOTECH GESELLSCHAFT FUER NEUE BIOTECHNOLOGISCHE PROZESSE UND
             PRODUKTE MBH
COMMENT   OS Hansenula polymorpha (yeast)
             PN JP 2002542788-A/1
             PD 17-DEC-2002
             PF 27-APR-2000 JP 200061405
             PR 27-APR-1999 DE 199 19 124.7
             PI GERD GELLISSEN,GERHARD BRAUS,RALPH PRIES,SVEN KRAPPMANN, PI
                 ALEXANDER W STRASSER
             PC C12N15/09,C07K16/40,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
                 ,C12N9/90,C12N15/00,
             PC C12N5/00
             CC Nucleic acid molecule, containing a nucleic acid which codes
             CC for a
             CC polypeptide with chorismate mutase activity
FEATURES   Location/Qualifiers
        source   1..843
                        /organism="Pichia angusta"
     FH   Key   Location/Qualifiers
     FT   source   1..843
     FT   /organism='Hansenula polymorpha (yeast)'
     FT   /mol_type="genomic DNA"
     FT   /db_xref="taxon:4905"
ORIGIN